

# A simple method to convert sink particles into stars

Mattia C. Sormani<sup>1</sup>, Robin G. Treß<sup>1</sup>, Ralf S. Klessen<sup>1,2</sup> and Simon C.O. Glover<sup>1</sup>

<sup>1</sup>*Institute for Theoretical Astrophysics, Zentrum für Astronomie der Universität Heidelberg, Albert-Überle-Str. 2, 69120 Heidelberg, Germany*

<sup>2</sup>*Universität Heidelberg, Interdisziplinäres Zentrum für Wissenschaftliches Rechnen, Im Neuenheimer Feld 205, D-69120 Heidelberg, Germany*

## ABSTRACT

Hydrodynamical simulations of star formation often do not possess the dynamic range needed to fully resolve the build-up of individual stars and star clusters, and thus have to resort to subgrid models. A popular way to do this is by introducing Lagrangian sink particles, which replace contracting high density regions at the point where the resolution limit is reached. A common problem then is how to assign fundamental stellar properties to sink particles, such as the distribution of stellar masses.

We present a new and simple statistical method to assign stellar contents to sink particles. Once the stellar content is specified, it can be used to determine a sink particle’s radiative output, supernovae rate or other feedback parameters that may be required in the calculations. Advantages of our method are (i) it is simple to implement, (ii) it guarantees that the obtained stellar populations are good samples of the initial mass function, (iii) it can easily deal with infalling mass accreted at later times, and (iv) it does not put restrictions on the sink particles’ masses in order to be used. The method works very well for sink particles that represent large star clusters and for which the stellar mass function is well sampled, but can also handle the transition to sink particles that represent a small number of stars.

**Key words:** stars: formation - stars: luminosity function - mass function - methods: numerical - methods: statistical

## 1 INTRODUCTION

Star formation involves the fragmentation of gaseous clouds, which then undergo local gravitational collapse. Following the global evolution of a cloud and, at the same time, the gravitational collapse of individual substructures is usually a task beyond the capabilities of modern numerical simulations (e.g. Klessen & Glover 2016).

A possible way to deal with this problem is the introduction of sub-resolution scale models for stellar birth. A very popular approach involves so-called sink particles.<sup>1</sup> This concept was first introduced by Bate et al. (1995), and then used and developed further by many authors (e.g. Krumholz et al. 2004; Jappsen et al. 2005; Federrath et al. 2010; Howard et al. 2014; Bleuler & Teyssier 2014; Klassen et al. 2016; Gatto et al. 2016, and references therein). In the sink particle approach, a contracting high-density region is replaced by a single Lagrangian particle at a stage where the numerical resolution limit in the simulation is reached. The particle inherits the mass as well as the linear and angular momentum of the original region, and in many implementations it can also accrete mass infalling at later times.

There are two regimes in which sink particles are used. If the resolution is high enough, each sink particle can represent a single star. If the resolution is poorer (for example because the simulated

region is very large), then a single sink particle may correspond to an entire star cluster. In the latter case, a rule is needed to assign to each sink particle the appropriate stellar content. Several algorithms have been proposed to perform this task (e.g. Howard et al. 2014; Dale et al. 2014; Gatto et al. 2016; Hu et al. 2016). However, they can usually be employed only if the sink particle mass distribution satisfies some restrictions. Typically they require a minimum mass to be exceeded, both at the time of formation and during the subsequent accretion phase. Moreover, even though they rely on stochastic sampling from the stellar initial mass function (IMF: see Kroupa 2002; Chabrier 2003) they can introduce biases which prevent the final result from being a completely faithful representation of the IMF.

In this short paper, we propose a new statistical method to assign a realistic stellar content to cluster sink particles. Once the distribution of stellar masses and formation times are determined, this information can be used to assign the radiative output of the sink particle, its supernovae rate, chemical yields, and so forth. The method is introduced in Section 2 and an illustrative example is presented in Section 3. We summarise and conclude in Section 4.

<sup>1</sup> sometimes also called star particles in situations where they are not allowed to accrete, usually in the context of cosmological simulations, e.g. Hopkins et al. (2014); Hu et al. (2016).

## 2 THE METHOD

### 2.1 Definition

Consider an IMF, with  $N$  different types of stars binned according to their masses. This defines a discrete set of stellar masses

$$\{m_1, m_2, \dots, m_N\}, \quad (1)$$

with corresponding mass fractions

$$\{f_1, f_2, \dots, f_N\} \quad \text{such that} \quad \sum_i f_i = 1, \quad (2)$$

i.e., in a well-sampled population a fraction  $f_i$  of the total mass will be in stars of type  $i$ . The statistical weights  $f_i$  can be easily determined starting from more common ways of parametrising the IMF (see Section 3) and depend on the width of the chosen bins.

Given a sink of mass  $M$ , we now describe our procedure to assign stars to it. Let us denote the stellar content by a vector

$$\mathbf{n} = \{n_1, n_2, \dots, n_N\}, \quad (3)$$

where the  $n_i$ 's are integers representing the number of stars of type  $i$ . So for example a sink with no stars has  $\mathbf{n} = \{n_1 = 0, n_2 = 0, \dots\}$  and a sink with just 2 stars of type 1 has  $\mathbf{n} = \{n_1 = 2, n_2 = 0, n_3 = 0, \dots\}$ . Let us call  $P(\mathbf{n})$  the probability that the stellar content assigned to the sink particle is  $\mathbf{n} = \{n_1, n_2, \dots\}$ . Our prescription is:

$$P(\mathbf{n}) = P_1(n_1)P_2(n_2)\dots P_N(n_N), \quad (4)$$

where

$$P_i(n_i) = e^{-\lambda_i} \frac{\lambda_i^{n_i}}{n_i!}, \quad (5)$$

and

$$\lambda_i = f_i \frac{M}{m_i}. \quad (6)$$

In other words, the number of stars of each type are assigned according to a Poisson distribution with mean  $\lambda_i$  which depends on the mass of the sink particle. The total stellar mass assigned to the sink will then be equal to

$$M_\star = n_1 m_1 + n_2 m_2 + \dots + n_N m_N, \quad (7)$$

and the stellar mass in stars of type  $i$  will be equal to:

$$M_{\star,i} = n_i m_i. \quad (8)$$

### 2.2 Properties

The method defined in the previous section has the following properties (see Appendix A for a derivation of these properties):

- (i) The total stellar mass is on average equal to the sink mass  $M$ :

$$\langle M_\star \rangle = \sum_{\mathbf{n}} M_\star P(\mathbf{n}) = M. \quad (9)$$

Here and in the following  $\langle \cdot \rangle$  denotes the ensemble average over many random realisations of the Poisson process and  $\sum_{\mathbf{n}} \equiv \sum_{n_1, n_2, \dots, n_N}$ .

- (ii) The variance of the total stellar mass  $M_\star$  is:

$$\frac{\langle (M_\star - M)^2 \rangle}{M^2} = \sum_{\mathbf{n}} \frac{(M_\star - M)^2}{M^2} P(\mathbf{n}) = \frac{\bar{m}}{M}, \quad (10)$$

where

$$\bar{m} = \sum_i f_i m_i \quad (11)$$

is the mean stellar mass according to the IMF.

- (iii) The average mass in stars of type  $i$  is:

$$\langle M_{\star,i} \rangle = \sum_{\mathbf{n}} m_i n_i P(\mathbf{n}) = f_i M. \quad (12)$$

- (iv) The variance of the mass in stars of type  $i$  is:

$$\alpha_i^2 \equiv \frac{\langle (M_{\star,i} - f_i M)^2 \rangle}{(f_i M)^2} = \sum_{\mathbf{n}} \frac{(m_i n_i - f_i M)^2}{(f_i M)^2} P(\mathbf{n}) = \frac{m_i}{f_i M}. \quad (13)$$

(v) The stellar content  $\mathbf{n}$  assigned to a sink particle of mass  $M = M_1 + M_2$  is statistically equivalent to the sum  $\tilde{\mathbf{n}} = \mathbf{n}_1 + \mathbf{n}_2$  of the stellar contents assigned to sink particles of masses  $M_1$  and  $M_2$ . In other words,  $\mathbf{n}$  and  $\tilde{\mathbf{n}}$  have exactly the same probability distribution. This is a direct consequence of the additive property of the Poissonian process, i.e. the sum of two Poisson distributions is a Poisson distribution whose mean is the sum of the means of the original distributions (e.g. Haight 1967).

### 2.3 Discussion of the properties

Since the method is stochastic in nature, a sink particle mass  $M$  and its stellar mass  $M_\star$  will in general differ. However property (i) ensures that given many realisations of the process the two masses are on average the same. The typical deviations of the stellar mass  $M_\star$  from the sink particle mass  $M$  are quantified by property (ii): they will be small if  $M \gg \bar{m}$ , which for typical applications is  $\bar{m} \sim$  few solar masses.

Property (iii) demonstrates that the IMF is always recovered on average. Property (iv) quantifies the deviation in the number of stars of each type obtained from the average number that would be expected according to the IMF. From equation (13) we see that the smaller the  $f_i$ 's (the rarer the star), the larger the deviations. The precise values of the  $f_i$ 's depend on how one chooses to discretise a continuous IMF by binning it. Broad bins will lead to smaller deviations, narrow bins will result in larger fluctuations. How one should choose the bin sizes depends on the fluctuations that can be tolerated and on the science question to be addressed. The method also works well with bins of unequal sizes. For example, if one is only interested in the number of high-mass stars, it may be appropriate to use fine binning in the high-mass regime, and combine all low-mass stars into one single bin. The stochastic fluctuations in each bin  $i$  depend on the sink particle mass and the bin width and can be calculated from equation (13). We discuss this further in Sect. 3.

Property (v) has nice consequences. First, accretion onto the sink particle is naturally taken care of. If a sink particle of mass  $M$  later accretes a mass  $\Delta M$ , the extra mass can be converted into stars and the result would be the same as if all the original plus the accreted mass is converted at once. Thus if the same sink particle accretes mass in many stages, each time we can convert the mass into stars without worrying that the final distribution of stars will be affected by when we do it. Moreover, if the sampling is done with sufficiently fine time stepping, the model will automatically produce a time-distribution for the birth of stars which can then be used in the scientific investigation at hand. If a sink with initial mass  $M \gg \bar{m}$  later accretes only a small quantity of mass (e.g., comparable to the mass of a star), then our method still works well. The extra mass will produce few, one or no stars, which will be assigned to the sink in addition to the stars assigned initially, when the sink was created. At this later time, the total stellar content of the sink will be a better sample of the IMF than it was before. So even if we accrete a small amount of mass, the overall content of

the sink always tends towards a better representation of the IMF, never worse.

Second, the only statistically important parameter for a collection of sink particles is their total mass  $M$ . The total distribution of stars is independent of how these masses are distributed across the sinks. This property makes our approach quite insensitive to the numerical resolution, because it does not matter if the mass of the emerging star cluster is contained within one single massive sink (in a lower resolution simulation) or distributed over particles of smaller masses (as would form in the same region in simulations with better spatial resolution). The overall distributions will be exactly the same.

The fact that the stellar mass and the sink mass can be different may create problems in certain circumstances where a careful conservation of mass and momentum is needed. The simplest way to address this issue is to let each particle have two masses: a dynamical mass ( $M$ ) and a stellar mass ( $M_*$ ). The code will continue to use  $M$  as the dynamical mass, while  $M_*$  is used only to calculate radiative output and stellar-related properties of the sink. In this way, the stellar content is just a label attached to the sink.

As discussed above, fluctuations will be small when  $M \gg \bar{m}$ . In the intermediate regime where, say,  $M \simeq 100M_\odot$ , property (ii) guarantees that fluctuations in the total mass will still be small. When approaching the low mass regime, in which the mass of an individual sink is comparable to the mass of an individual star,  $M \simeq \bar{m}$ , our scheme can still be used but one needs to evaluate the situation more carefully. In this regime, our scheme will usually assign few, one or no stars at all to each sink. Clearly, the same happens in nature for small clusters which contain only few stars. Hence, our scheme automatically covers the fluctuations of the IMF that we do see in real clusters of small mass. Obviously both in our scheme and in nature the stellar content of an individual sink or small cluster containing very few stars cannot be a good sample of the IMF. Even if the stellar content of individual sinks can fluctuate wildly, when we consider many small clusters of small mass together, we eventually recover the IMF (thanks to property (v) above), exactly as it would happen in nature if we did a statistical study by considering many small clusters together.

However our scheme has a shortcoming and allows for something that nature forbids: the stellar mass  $M_*$  assigned to a sink can be larger than the sink mass  $M$ . This effect is much more severe in the regime of small sink masses (see equation 10). Whether this is a problem in practice depends on the aims of the scientific investigation at hand. For example, in feedback simulations in which one is interested in producing a realistic population of supernovae and analysing their effect on the surrounding ISM, our scheme works well. It can produce a realistic distribution of supernovae, both in space and in time. Our method works less well if we are interested in modelling the evolution of one particular cluster (rather than the average effect of many clusters on the ISM through their stellar feedback), as in this case we may assign to the sink of interest a stellar mass  $M_*$  which is too far off from its mass  $M$ .

Finally, we note that depending on the underlying physical model not all the mass in the sink particle needs to be converted into stars, i.e. the star formation efficiency may not be 100%. In this case, we should modify our procedure by replacing  $M$  by an effective mass  $M_{\text{eff}}$  in the definition of  $\lambda_i$  in equation (6). For example, if the star formation efficiency is, say, 50%, then we can use  $M_{\text{eff}} = 0.5M$ .

## 2.4 Comparison with other methods

Other methods that have been proposed (e.g. Howard et al. 2014; Gatto et al. 2016) usually involve a procedure along the following lines. First, a gas reservoir is associated to the sink particle. The reservoir is then gradually converted into stars by sampling directly from the IMF. At each step of the sampling process, the gas reservoir diminishes. The main problem with such methods is to decide at which point (i.e., for which reservoir mass) one should stop sampling. For example, when the gas reservoir is too small, sampling from the IMF can lead to a star whose mass exceeds that available in the reservoir. What decision should be made at this point?

There are various possibilities. Howard et al. (2014) accepts the outcome of the sampling only if the mass of the star does not exceed the mass of the reservoir. However, this introduces a bias and the IMF will be oversampled at the low-mass end. To circumvent this problem, these authors assumed that accretion is present at a certain rate and devise a procedure for fine-tuning the sampling time, and have shown that under such conditions the bias is negligible. However this introduces restrictions on the sink particles mass distribution in order to work and depends on the details of the accretion scheme and in general on the details of the simulations.

Another possibility is to stop the sampling process when the gas reservoir is too small. However this method cannot handle sink particles whose mass is below that of the most massive star. Gatto et al. (2016) and Hennebelle & Iffrig (2014) follow an approach along similar lines. These authors are mostly concerned about the high-mass end of the IMF ( $M > 8M_\odot$ ), which is needed to model stellar feedback in the ISM. They randomly draw one massive star for every  $120M_\odot$  of accreted mass. The difference between this value and the sampled mass is assumed to go into low-mass stars. Since the mass of each sink is not in general an exact multiple of  $120M_\odot$ , this produces an overall deficit of high mass stars and the procedure has to be modified by considering the mass of all the sinks together (Körtgen et al. 2016). The stars so created have then to be redistributed to the sinks according to some criteria.

Simpler methods that do not involve statistical sampling from the IMF have also been proposed to assign specific properties to sink particles. Dale et al. (2014), for example, compute the mass fraction in high mass stars but then turn on the corresponding feedback only if the sink mass is above a certain threshold value. Hopkins et al. (2014) and Hu et al. (2016) assign stochastically discrete forms of feedback, such as supernovae, to the sinks based on their mass, age and metallicity. The limitations of these models are linked to the lack of a full stellar content of the sink, and so the properties inferred this way can only be coarse-grained representations of the true stellar population. This may be appropriate when modelling stellar feedback in cosmological galaxy formation calculations, where individual sink particles are very massive and can contain multiple clusters, but it is not sufficient in high resolution simulations where sink particles represent small stellar clusters or in situations where a more detailed modelling of stellar evolution at the sub-grid level is needed.

Our method bypasses most of the problems mentioned above. It guarantees that the IMF is well sampled without putting any restriction on the sink particle masses or on the sampling time. It also makes it easy to control the average stellar mass produced, so that unwanted overall deficits or surpluses are avoided. It naturally gives the full stellar content, including the stars of each type, and not only the number of high mass stars. Moreover, our method is conceptually simple and computationally efficient since it only

needs to draw from independent Poisson distributions for which standard routines are available in all programming languages.

As discussed in Section 2.3, a possible shortcoming of our method is that it is in principle possible to assign to a sink a stellar mass  $M_*$  that exceeds the mass of the sink  $M$ . However, this is unlikely to be a problem in most applications, particularly when the star formation efficiency is less than 100% and  $M \gg \bar{m} \simeq \text{few } M_\odot$ , in which case the variance can also be seen as a fluctuation in the star formation efficiency. As mentioned in Sect. 2.3, we suggest treating the stellar content as a mere label attached to the particle, and keeping the sink particle mass as the dynamical mass in the underlying hydrodynamic simulation.

### 3 EXAMPLE WITH A KROUPA IMF

In this section we briefly apply our method with the IMF as parametrised by Kroupa (2002). In this approach, the number of stars in the mass interval  $(m, m + dm)$  is given by  $f(m)dm$  with

$$f(m) = \begin{cases} Am^{-0.3} & \text{for } \tilde{m}_1 \leq m < \tilde{m}_2 \\ Ak_1 m^{-1.3} & \text{for } \tilde{m}_2 \leq m < \tilde{m}_3 \\ Ak_2 m^{-2.3} & \text{for } \tilde{m}_3 \leq m \end{cases}, \quad (14)$$

and where  $A$  is a global normalization factor,  $k_1 = \tilde{m}_1^{-0.3+1.3}$ ,  $k_2 = k_1 \tilde{m}_2^{-1.3+2.3}$ ,  $\tilde{m}_1 = 0.01 M_\odot$ ,  $\tilde{m}_2 = 0.08 M_\odot$  and  $\tilde{m}_3 = 0.5 M_\odot$ .

We have binned Kroupa's IMF in mass bins in the range  $M_{\min} = 0.01 M_\odot$  to  $M_{\max} = 100 M_\odot$ . The  $m_i$ 's and  $f_i$ 's are calculated according to the following formulae:

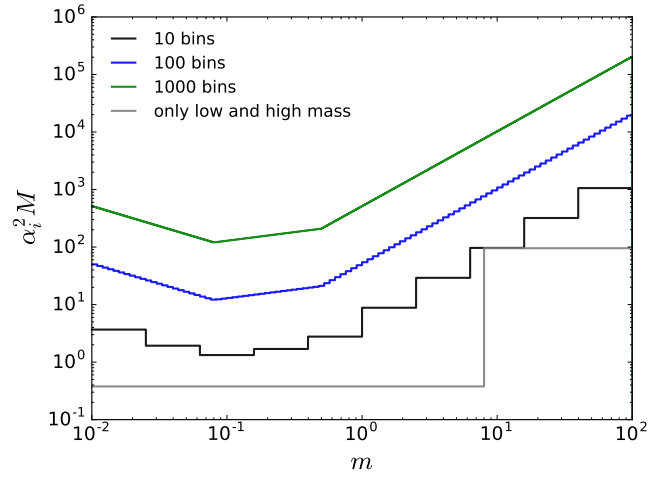
$$m_i = \frac{\int_i m f(m) dm}{\int_i f(m) dm}, \quad (15)$$

$$f_i = \frac{\int_i m f(m) dm}{\int_{M_{\min}}^{M_{\max}} m f(m) dm}, \quad (16)$$

where  $\int_i$  denotes the integral over the  $i$ -th bin.

Figure 1 shows the value of  $\alpha_i^2 M$  for four different ways of binning. The quantity  $\alpha_i$  is the expected error in the number of stars assigned to each bin, i.e., it is the deviation from the number actually assigned in a random realisation and the average number predicted by the IMF (see equation 13). We can use Fig. 1 to read off directly the minimum mass  $M$  needed for the error  $\alpha_i$  to be below a given tolerance value. The three top lines in Fig. 1 refer to logarithmic binnings with 1000, 100 or 10 bins, respectively. We see that, as one would expect, for a particle of given mass  $M$  fine binning leads to larger fluctuations. This means that if we want to distinguish between many types of stars and still get a well sampled IMF, we need sink particles with a large mass that can contain many stars (the same is true in nature for real clusters: we cannot have a good sample of the IMF looking at a too small cluster). The gray line at the bottom is for a very coarse-grained binning that distinguishes only between low mass ( $< 8 M_\odot$ ) and high mass ( $> 8 M_\odot$ ) stars, and it shows that in this case lower values of  $M$  are permitted to achieve a given tolerance target.

Each panel in Fig. 2 shows a random realisation calculated according to our method for a sink particle of given mass  $M$  and a binning with 100 logarithmic bins (blue line in Fig. 1). Different panels differ only in the value of  $M$ . It is clear that when  $M = 1 M_\odot$  the IMF is not well sampled. As  $M$  is increased, the lower-mass end of the IMF gets well sampled first, while the higher-mass region retains larger sampling uncertainties, as one would expect from the steeply decreasing probability of forming high-mass stars. Figure 1 also explains this behaviour. For the case of 100 logarithmic bins,



**Figure 1.** The values of  $\alpha_i^2 M$ , which quantifies the fluctuations in the number of stars in each bin, for four different binnings of the IMF in the range  $0.01 M_\odot$ - $100 M_\odot$ . The first three lines from the top are for logarithmic binnings with 1000, 100 or 10 bins respectively. The bottom line is for a binning that distinguishes only between low mass ( $< 8 M_\odot$ ) and high mass ( $> 8 M_\odot$ ) stars.

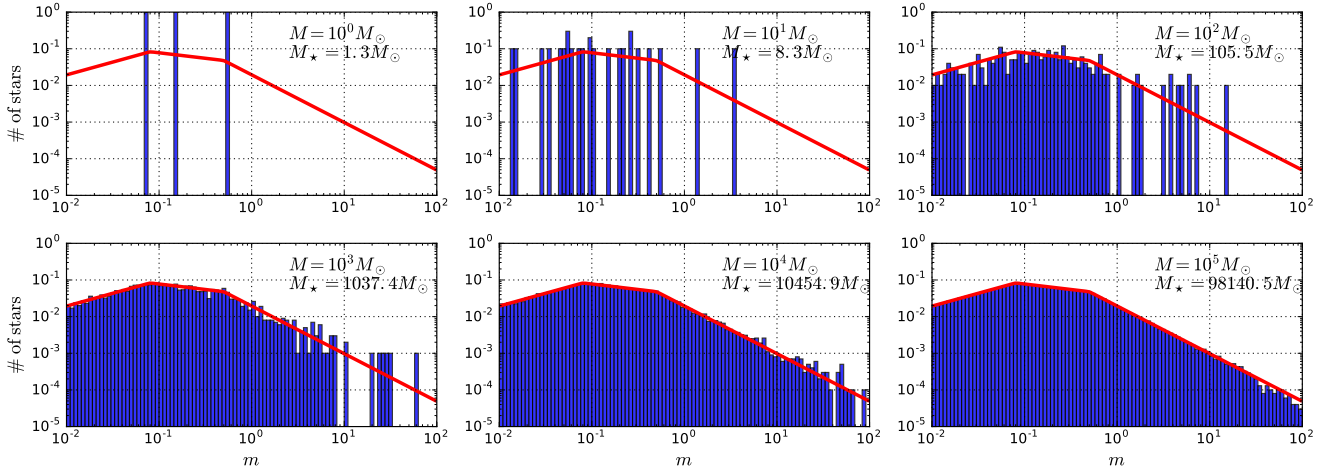
the blue line indicates that the fluctuations at the low-mass end are small (say  $\alpha_i \lesssim 10\%$ ) for  $M \gtrsim \text{few } 10^3 M_\odot$ , and at the high-mass end are small when  $M \gtrsim \text{few } 10^4 M_\odot$ , which is exactly the behaviour seen in Fig. 2. If we can get by with fewer mass bins, we can reach the same fluctuation level  $\alpha_i$  with lower-mass sinks. In many applications, the decision on how much variance  $\alpha_i$  can be tolerated at a certain mass  $m_i$  determines the choice of the bin size  $f_i$ , which can then be calculated from equation (13) for a given mass  $M$ . Keeping in mind property (v), we can interpret Figs. 2 and 1 either as referring to the mass of a population of sink particles summing up to a total of  $M$  or as referring to one single sink particle with mass  $M$ . This means we have the choice of whether we apply the maximum tolerance to a single sink particle or to the population as a whole.

We note that in this example we have focussed on the Kroupa (2002) IMF and discretised it in four different ways. However, the method can be applied with no extra complication to different IMF models and to bins of any sizes and non-uniform width. All one needs to do is calculate the appropriate values of the  $f_i$ 's and  $m_i$ 's for the desired IMF and mass binning. The method can be also generalised to the continuous limit if one wants to avoid binning (see Appendix B).

### 4 CONCLUSION

We have presented a simple prescription for assigning stellar contents to sink particles used in hydrodynamical simulations with star formation that do not fully resolve the build-up of individual stars and star clusters. The assigned stars can then be used to determine the sink particle's radiative output, its supernova rate and chemical yield, or other parameters of interest related to the underlying stellar population. The key properties of our method are summarised as follows:

- It guarantees that the assigned stellar populations are a faithful representation of the IMF.



**Figure 2.** Application of our method using Kroupa’s IMF. Different panels show random realisations for different values of the mass  $M$ . The blue bars show the number of stars in each mass bin, normalised with arbitrary scaling. The red line shows the number of stars in each bin that would be expected according to Kroupa’s IMF.

- It can easily deal with infalling mass accreted at later times by the sink particle.
- It does not put any restriction on the sink particle masses or on the sampling time. For example, there is no minimum mass requirement for the sink particles. The method therefore is relatively insensitive to the numerical resolution adopted in the simulation.
- It can accommodate in a straightforward way non-uniform mass bins and arbitrary choices of the IMF.
- It can deal with sinks of intermediate masses, i.e. in the intermediate regime between the cluster regime and the star regime, thus providing a smooth transition between the two.
- In the intermediate-small sink mass regime, it can cover the statistical fluctuations seen in nature in real clusters of small size.
- It is computationally very efficient and easy to implement.

Some possible shortcomings are:

- Due to the stochastic nature of the method, the sink particle mass  $M$  will be equal to its stellar mass  $M_*$  only on average, but not in every single instance. This can cause problems when approaching a physical regime where sink particles describe few or single stars. Our method works best when  $M \gg \bar{m}$  as this shortcoming becomes negligible in this limit.
- The IMF is taken as input to the model. It is independent of the sink particles’ mass distribution from the numerical simulation. That means that this quantity cannot be used as a diagnostic tool.

## ACKNOWLEDGEMENTS

The authors thank Paul Clark, Samuel Geen, Diederik Kruijssen, Mordecai Mac Low, John Magorrian, Eric Pellegrini, Rowan Smith and all the members of the Heidelberg Institute for Theoretical Astrophysics for useful comments. We also acknowledge support from the Deutsche Forschungsgemeinschaft in the Collaborative Research Center (SFB 881) “The Milky Way System” (subprojects B1, B2, and B8) and in the Priority Program SPP 1573 “Physics of the Interstellar Medium” (grant numbers KL 1358/18.1, KL 1358/19.2). RSK furthermore thanks the European Research Council for funding in the ERC Advanced Grant STARLIGHT (project number 339177).

## APPENDIX A: DERIVATION OF THE PROPERTIES

In this appendix we prove the properties listed in Section 2.2. It is well known that the Poisson distribution given by equation (5) has the following properties:

$$\sum_{\mathbf{n}} P(\mathbf{n}) = 1, \quad (\text{A1})$$

$$\sum_{\mathbf{n}} n_i P(\mathbf{n}) = \lambda_i, \quad (\text{A2})$$

$$\sum_{\mathbf{n}} n_i^2 P(\mathbf{n}) = \lambda_i + \lambda_i^2. \quad (\text{A3})$$

Properties (i) and (iii) in Section 2.2 follow immediately from equations (A1), (A2) and (A3). Property (ii) can be proved as follows

$$\frac{\langle (M_* - M)^2 \rangle}{M^2} = \frac{1}{M^2} \sum_{\mathbf{n}} M_*^2 P(\mathbf{n}) - 1 \quad (\text{A4})$$

$$= \frac{1}{M^2} \sum_{\mathbf{n}} (n_1 m_1 + \dots + n_N m_N)^2 P(\mathbf{n}) - 1 \quad (\text{A5})$$

$$= \frac{1}{M^2} \left[ \sum_i \sum_{\mathbf{n}} n_i^2 m_i^2 P(\mathbf{n}) + 2 \sum_{i < j} \sum_{\mathbf{n}} n_i n_j m_i m_j P(\mathbf{n}) \right] - 1 \quad (\text{A6})$$

$$= \frac{1}{M^2} \left[ \sum_i m_i^2 (\lambda_i + \lambda_i^2) + 2 \sum_{i < j} m_i m_j \lambda_i \lambda_j \right] - 1 \quad (\text{A7})$$

$$= \frac{1}{M^2} \left[ \sum_i m_i f_i M + f_i^2 M^2 + 2 \sum_{i < j} f_i f_j M^2 \right] - 1 \quad (\text{A8})$$

$$= \frac{1}{M^2} \left[ \sum_i m_i f_i M + \left( \sum_i f_i \right)^2 M^2 \right] - 1 \quad (\text{A9})$$

$$= \frac{1}{M^2} \left[ \sum_i m_i f_i M + M^2 \right] - 1 \quad (\text{A10})$$

$$= \frac{\bar{m}}{M} \quad (\text{A11})$$

which is equation (10). Property (iv) can be proved as follows:

$$\alpha_i^2 = \frac{\langle (M_{*,i} - f_i M)^2 \rangle}{(f_i M)^2} \quad (\text{A12})$$

$$= \sum_{\mathbf{n}} \frac{(m_i n_i - f_i M)^2}{(f_i M)^2} P(\mathbf{n}) \quad (\text{A13})$$

$$= \sum_{\mathbf{n}} \frac{m_i^2 n_i^2 + f_i^2 M^2 - 2m_i n_i f_i M}{(f_i M)^2} P(\mathbf{n}) \quad (\text{A14})$$

$$= \frac{m_i^2 (\lambda_i + \lambda_i^2) + f_i^2 M^2 - 2m_i \lambda_i f_i M}{(f_i M)^2} \quad (\text{A15})$$

$$= \frac{m_i f_i M + f_i^2 M^2 + f_i^2 M^2 - 2f_i^2 M^2}{(f_i M)^2} \quad (\text{A16})$$

$$= \frac{m_i}{f_i M} \quad (\text{A17})$$

which is equation (13). Property (v) follows immediately from the additive property of Poisson distributions quoted in the main text.

## APPENDIX B: CONTINUOUS LIMIT

Our method involves discretisation of the IMF, which may not be ideal in certain applications where one wants to avoid binning and deal directly with a continuous IMF. In the continuous limit, when the number of bins tends to infinity and the width of each bin tends to zero, our method gives results that are equivalent to the following two-step procedure. Given a sink particle of mass  $M$ :

- (i) Draw the number of stars to assign to the sink according to a Poisson distribution with mean  $\lambda = M/\langle m \rangle$ .
- (ii) For each star, draw its mass directly from the (continuous) IMF.

Only for this appendix, we have defined

$$\langle m \rangle = \frac{\int m f(m) dm}{\int f(m) dm}, \quad (\text{B1})$$

where  $f(m)$  is the IMF (see Section 3).

## REFERENCES

- Bate M. R., Bonnell I. A., Price N. M., 1995, MNRAS, 277, 362  
 Bleuler A., Teyssier R., 2014, MNRAS, 445, 4015  
 Chabrier G., 2003, PASP, 115, 763  
 Dale J. E., Ngoumou J., Ercolano B., Bonnell I. A., 2014, MNRAS, 442, 694  
 Federrath C., Banerjee R., Clark P. C., Klessen R. S., 2010, ApJ, 713, 269  
 Gatto A. et al., 2016, ArXiv e-prints, 1606.05346  
 Haight F., 1967, Handbook of the Poisson distribution, Publications in operations research. Wiley  
 Hennebelle P., Iffrig O., 2014, A&A, 570, A81  
 Hopkins P. F., Kereš D., Oñorbe J., Faucher-Giguère C.-A., Quataert E., Murray N., Bullock J. S., 2014, MNRAS, 445, 581  
 Howard C. S., Pudritz R. E., Harris W. E., 2014, MNRAS, 438, 1305  
 Hu C.-Y., Naab T., Walch S., Glover S. C. O., Clark P. C., 2016, MNRAS, 458, 3528  
 Jappsen A.-K., Klessen R. S., Larson R. B., Li Y., Mac Low M.-M., 2005, A&A, 435, 611

- Klassen M., Pudritz R. E., Kuiper R., Peters T., Banerjee R., 2016, ApJ, 823, 28  
 Klessen R. S., Glover S. C. O., 2016, in: Star Formation in Galaxy Evolution: Connecting Numerical Models to Reality, Saas-Fee Advanced Course, 43, 85  
 Körtgen B., Seifried D., Banerjee R., Vázquez-Semadeni E., Zamora-Avilés M., 2016, MNRAS, 459, 3460  
 Kroupa P., 2002, Science, 295, 82  
 Krumholz M. R., McKee C. F., Klein R. I., 2004, ApJ, 611, 399